Ascertaining Mathematical Theorems

Citation for published version:
McCasland, R, Bundy, A & Smith, P 2006, 'Ascertaining Mathematical Theorems', *Electronic Notes in Theoretical Computer Science*, vol. 151, no. 1, pp. 21-38. https://doi.org/10.1016/j.entcs.2005.11.021

Digital Object Identifier (DOI):
10.1016/j.entcs.2005.11.021

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Electronic Notes in Theoretical Computer Science

Calculemus 2005 Preliminary Version

Ascertaining Mathematical Theorems [1]

Roy L. McCasland [a,2], Alan Bundy [a,3] and Patrick F. Smith [b,4]

[a] *School of Informatics*
University of Edinburgh
Edinburgh, Scotland, UK

[b] *Department of Mathematics*
University of Glasgow
Glasgow, Scotland, UK

Abstract

Whereas to most logicians, the word "theorem" refers to any statement which has been shown to be true, to mathematicians, the word "Theorem" is, relatively speaking, rarely applied, and denotes something far more special. In this paper, we examine some of the underlying reasons behind this difference in terminology, and we show how this discrepancy might be exploited, in order to build a computer system which automatically selects the latter type of "Theorems" from amongst the former. Indeed, we have begun building the automated discovery system MATHsAiD, the design of which is based upon our research. We provide some preliminary results produced by this system, and compare these results to Theorems appearing in various mathematics textbooks.

Key words: automated theorem generation, mathematical reasoning

1 Introduction

Whereas to most logicians, the word "theorem" refers to any statement which has been shown to be true, to mathematicians, the word "Theorem" [5] is, relatively speaking, rarely applied, and denotes something far more special.

---

[1] The authors are supported by EPSRC MathFIT grant GR/S31099.

[2] Email: rmccasla@inf.ed.ac.uk
[3] Email: A.Bundy@ed.ac.uk
[4] Email: pfs@maths.gla.ac.uk

[5] Throughout this work, "theorems" will be used to denote truths in the logicians' sense, whereas "Theorems" will denote those truths which mathematicians would typically consider worth recording, including lemmas, corollaries, propositions, etc. Of course, some allowances should be made for differences of opinion.

This is a preliminary version. The final version will be published in
Electronic Notes in Theoretical Computer Science
URL: www.elsevier.nl/locate/entcs

In this paper, we examine some of the underlying reasons behind this difference in terminology, and we show how this discrepancy might be exploited, in order to build a computer system which automatically selects the "Theorems" from amongst the "theorems".

The importance of this distinction should not be underestimated, since it would seem that the ability on the part of mathematicians to 'separate the wheat from the chaff' (in their view) plays a major role in enabling them to continue making new discoveries, while not recording so many theorems that they are unable to cope with all the data. It seems reasonable to expect that, should a machine learning or an artificial intelligence system be able to mimic, at least to some extent, this same ability, then this system might also be able to continue "discovering" new Theorems, avoiding the usual scaling-up difficulties. In particular, such a system could quite conceivably prove very useful in certain applications (e.g., formal methods), where at present, it is difficult to dispatch all the proof obligations that arise, without first finding and proving some necessary (or, at least, helpful) Lemmas. In any event, short of finding some means of identifying those truths which are in some sense *new*, the alternatives for any such system would be either to store every newly discovered "theorem" (in the logicians' sense), or to store none – in effect, either memorize everything, or learn nothing.

We have, in fact, begun building a computer system, called **MATHsAiD** (**M**echanically **A**scertaining **T**heorems from **H**ypothese**s**, **A**x**i**oms and **D**efinitions), with the main goal being to automatically discover and identify truths that mathematicians would call Theorems. Although this work is still in relatively early stages, the results are quite promising. The reader should understand, though, that this paper is not intended to provide a system-description – nor is it intended to explain in detail our methods of generating [6] the theorems, from which the Theorems are chosen. Rather, at present, we concentrate on the principles and the philosophy behind the Theorem-filtering part of the system. We do, however, include in the Appendix a sample of statements that MATHsAiD has selected as Theorems. It should perhaps be noted that our research has benefited from the fact that two of the authors have had considerable experience in mathematical research, and another author, who was trained as a logician, has had considerable experience in artificial intelligence research. It should also be noted that our ideas have been influenced, in no small part, by the work expressed in [4] and [6].

The authors are aware of others who have done somewhat related work

[6] Roughly speaking, whenever MATHsAiD is given a collection of axioms (definitions) to investigate, it sets up a sequence of sets of hypotheses, and for each set of hypotheses, in turn, it uses a forward-chaining process to generate theorems that follow from these hypotheses – bearing in mind the criteria described in section 3. The sequence is determined by the nature of each of the axioms. That is to say, the axioms are divided into classes (e.g., relations, operations, closure-type, etc.), and each class is treated in such a way as to (hopefully, at least) enable MATHsAiD to find the more routine results one might look for, in regards to said class.

(see, for example, [2], [3], [5], [7] and [11]). However, to our knowledge, no one else has taken our approach to this particular problem, nor has anyone achieved entirely satisfactory results.

2 Some Alternative Approaches

Before moving on, we briefly mention some alternative approaches, which we and others[7] have considered. One could attempt to apply various heuristics, the most popular of which seems to be "interestingness". However, we believe that trying to quantify "interestingness" is neither practical nor useful, since this notion, at least as it is applied to mathematical Theorems, is entirely relative.[8] Not only will there be some disagreement amongst various mathematicians in this regard, but what is interesting to a particular mathematician today, will not necessarily be of interest to him (or her) a year hence. At any rate, most of the Theorems, as recorded in textbooks, would likely not score terribly well, as judged by mathematicians, in the "interestingness" category. Nevertheless, mathematicians would, as a rule, agree that the same Theorems are appropriately labelled as such.

Alternatively, one could attempt to measure "importance" or "usefulness". For example, one could count the number of times a given theorem is used in proving successive theorems, and then select as Theorems only those truths whose count exceeds some threshold. One can even expand this further, by attaching weights to each of the successive theorems, and adjusting the count for the given theorem accordingly. At first glance, this approach might seem somewhat promising. However, many of the tools which mathematicians find most useful in their research, would not, in fact, be considered by them to be Theorems. Moreover, the outcome would depend heavily upon the characteristics of the theorem prover used, and upon the order in which the material is introduced. We therefore expect that the results of this approach would be less than satisfactory.

In fact, from our point of view, the main drawback to this approach – and indeed, the approach used by most other existing systems – is that it in no way captures the human mathematical process[9], and is therefore contrary to our basic philosophy (see section 3). This, in fact, is the primary difference between our system and the others already mentioned.

[7] See the references towards the end of the preceeding section.

[8] Indeed, the whole notion as to what precisely constitutes a mathematical Theorem is, to some degree, relative. There is, in fact, no "gold standard" against which we might measure our results. That said, by comparing our results with mathematics textbooks (see section 4), we can get some sense as to whether our efforts have met with at least some success.

[9] In particular, other systems rely on existing ATP's to provide many of the measures used in determining "interestingness", "usefulness", "novelty", etc. Effective and efficient as ATP's are, they do not capture the mathematical process, as practiced by humans. In particular, no attempt is made by other systems to distinguish between "mathematics" and "logic".

3 Mathematics and Logic

Our basic philosophy, regarding MATHsAiD, is that all methods used in the system should, in some sense, preserve the human mathematical process, insofar as it is both possible and prudent. We believe that, by and large, the more closely we adhere to the human process, the better the results – and the more likely that such desirable traits (e.g., "importance", "usefulness", "novelty", etc.) will be found amongst the results, without our having to either search for, or measure them directly.

For this paper, we focus on identifying how mathematicians determine which truths to call Theorems, and which to set aside – true though they may be. In this regard, we have found four main properties which help mathematicians to do just that. The first is that every Theorem should satisfy the property of being, at least in some sense, *new*. The next two properties work in tandem; each Theorem should either be as *simple* as possible, or else provide a *sharper bound* than any previously discovered Theorem. The last property stems from the fact that mathematicians tend to prefer logical equivalences, whenever they can get them – subject to the first property, that is. Therefore, for each Theorem identified by MATHsAiD, as well as for each axiom given, the system tries to prove the *converse*, provided that the converse is in some sense a substantive statement.

It is fair to say that these ideas are all rather simple. This, in our view, is one of the main strengths of our approach, since quite often, it is the simplest ideas which produce the best results.

3.1 Already Known vs Newly Discovered

Consider again the fact that, to most logicians, every proven statement should be called a theorem. This suggests that, insofar as their thinking is concerned, every true statement is deemed to be as important (or as unimportant) as every other true statement. Were this not the case, then one would expect the nomenclature to differentiate amongst the various types of truths.

On the other hand, mathematicians have lots of different names for what they perceive to be different sorts of truths (e.g., Lemma, Theorem [10], Corollary), suggesting that mathematicians not only want to make a distinction as to the importance of various truths, but also to provide a sense of the history in the discovery of those truths. That is, a Lemma [11] is typically so-called because it has been useful in proving a Theorem, and a Corollary typically follows easily from a Theorem. Crucially, however, most theorems are considered by mathematicians to be unworthy of any sort of special name.

Practically every mathematician has either read, heard, or spoken the

[10] In this paragraph, the word Theorem refers to statements given this specific title by mathematicians – as opposed to our meaning of this word throughout the rest of this work.
[11] It must be said that logicians also refer to Lemmas.

phrase, "It can easily be shown that...", or the statement, "This result is already known". While each of these lines is sometimes used in a rather condescending manner, the truth behind them helps to underline the difference between the logician's and the mathematician's points of view.

Of course, to a mathematician, "already known", does not, by any stretch, simply apply to those statements which have already been published. Consider the reputation amongst mathematicians for being able, upon looking at a Theorem, to fairly quickly infer several other conclusions – more so than would likely be apparent to the uninitiated. It follows that, to a mathematician, once a Theorem has been "newly discovered", then many other theorems immediately become "known" as well. For this reason, these other theorems are not considered to be new Theorems [12].

Purely for the sake of convenience, let us assume that the notion of "already known" is well-defined. One could then build an ascending chain of sets K_t of statements which are understood to be already known at time t, with the initial state K_0 being the set consisting of the axioms, along with the rules of logic. As already suggested, from a mathematician's perspective, when a theorem s has newly become known at time t (thus $s \notin K_t$), then there exists a set S_t (to which s belongs) of statements that have now likewise become "newly known". As such, then, we have $K_{t+1} = K_t \cup S_t$. It should be noted that mathematicians would not necessarily declare s, or for that matter, any other member of S_t, to be a Theorem.

Admittedly, the notion of "already known" is not at all necessarily well-defined. Not only is the choice of s wide open, but there would almost certainly be at least some disagreement amongst mathematicians about the contents of K_{t+1}, even if there were agreement on the choice of s. That said, mathematicians would most likely agree that the cardinality of S_t is almost always greater than 1, whereas logicians would probably say that the cardinality of S_t is always 1.

In any event, in practical terms, we do not need to, nor do we want to actually compute either of the sets K_t or S_t, for any value t. Our main priority is to make certain, for any statement s which we might consider calling a Theorem, that s is not "already known" [13] – however vague that notion might be. We are, after all, only attempting to find a sort of 'best approximation' to the human mathematical process.

3.2 Mathematics or Logic

Both mathematicians and logicians agree that every step in a proof must be logically sound, and likewise they generally agree on the types of arguments

[12] Aside from the occasional Corollary, that is.

[13] As an example of how this is currently implemented in MATHsAiD, once the commutativity and associativity of intersection is known, then so too is any simple combination of the two properties – e.g., $A \cap (B \cap C) = C \cap (A \cap B)$.

allowed in a proof. Indeed, both groups make certain distinctions amongst the sort of proof steps used, though the distinctions made by mathematicians would likely be rather coarser than those made by logicians. That said, both groups accept that there is a difference between 'logical' and 'mathematical' proof steps.

The difference lies in what one does with these distinctions. A mathematician tends to view logic merely as one of several tools – an indispensable tool, no doubt – but a tool, nonetheless. Rather like an astronomer with a telescope, who would rather look through the telescope than at it, a mathematician prefers to study mathematics, and logic provides a means of doing just that.

Thus, from their vantage point, mathematicians perceive not only the axioms, the logic, and the (already discovered) Theorems to be "already known", but also any statements which, in their view, are trivially derived from any of these. By 'trivially derived', we mean, roughly speaking, anything that can be proven, using only (what mathematicians think of as) logic steps. Thus, in the notation above, once a statement s has been proven, then S_t would contain all statements trivially derivable from s, perhaps in combination with any members of K_t. Or in other words, for mathematicians to consider a statement to be new (and therefore, a potential candidate as a Theorem), its proof must contain something that they can identify as mathematics.

Therefore, in MATHsAiD, we attempt to distinguish (as mathematicians would) between 'logical' and 'mathematical' proof steps. In order, then, for any proven statement to be reported as a Theorem, it's proof must contain at least one 'mathematical' step. Of course, this is a necessary,[14] but not a sufficient condition.

To take a very simple example, we again consider the commutativity of intersection. Obviously, this property is a consequence of the commutativity of the logical 'and'. However, to most mathematicians[15], the commutativity of 'and' is not a Theorem (it is only logical!). On the other hand, the commutativity of the mathematical concept of intersection is a Theorem, because neither the intersect axiom nor the commutativity of 'and' tells us explicitly that intersection commutes; we must have the mathematical step of converting from 'and' to 'intersect'.

3.3 Simplicity

Mathematicians are notorious for always wanting statements to be in 'simplest' form. We believe that this tendency on their part not only derives from a deep

[14] There is one exception to this rule; whenever a converse of either an axiom or a Theorem is true, then this converse is also reported as a Theorem. See section 3.5.

[15] Mathematical logicians would more than likely disagree. But again, it depends upon what one is trying to achieve. After all, one's perception depends, to a considerable degree, upon one's focus.

sense of the nature of reality – physicists and mathematicians alike generally hold the tenet that the simplest explanations are the ones closest to the Truth – but from practical considerations as well. They find it quite helpful to assign names to certain concepts, in particular the more useful ones. In effect, by providing an equivalent, but simpler form, they are telling one another what they think is important.

They are also, as it happens, giving us a means of identifying Theorems – namely, simplicity. Within a class of equivalent statements, it is usually the simplest one which mathematicians choose – if any one is chosen – to record as a Theorem. There are, of course, exceptions to this rule, and we discuss one of these exceptions in the next section.

However, it is not just the simplest among a collection of equivalent statements that seems special to mathematicians. During the discovery process, whilst following a particular line of reasoning, a mathematician might prove many statements, perhaps few of which are equivalent to each other. The resulting sequence of theorems can vary substantially, with regards to the simplicity of each. Again, within this sequence, it is usually the simplest statement, if any, that a mathematician chooses to call a Theorem.

Not surprisingly, then, in MATHsAiD, we consider (as a rule) only the simplest statements as potential theorems. The simplicity comparison is made between equivalent statements, and between statements, one of whose proof list [16] is contained within the other's. This latter comparison represents the line of reasoning mentioned above. The only exceptions to this simplicity rule, thus far, at least, are discussed in the next two sections.

Definition 3.1 The *simplicity measure* of a mathematical statement S is given by the sum of the measures of the functors of S, where every functor measures one, except 'equals', which has measure one-tenth, and quantifiers, each of which has measure one-fourth.

To illustrate the simplicity comparison, [17] we provide two examples. [18] It turns out that when MATHsAiD finds the result $A \cap A = A$ (given that A is a set), it also finds that $A \cap (A \cap A) = A$. The latter result of course follows from the former, and indeed, its proof list contains the former's. One might suppose that this is sufficient reason for its being discarded, rather than for failing the simplicity test. The next example shows that this is not the case. Given that $a, b \in G$ (where G is a group) and $b * a = e$, MATHsAiD finds, in succession, that a^{-1} is equal to each of the terms $e * a^{-1}$, $(b * a) * a^{-1}$, $b * (a * a^{-1})$, and $b * e$. Finally, it finds that $a^{-1} = b$. The proof list of this last equality thus contains the proof list for each of the preceeding equalities, but of course, the last equation is the simplest.

[16] A list that records all the steps involved in the proof of the given statement.
[17] Note that no regard is given to the arity of any functors. We have, in fact, tried using more complicated measures. However, this relatively simple measure seems to work best.
[18] The corresponding Theorems can be found in the Appendix.

3.4 Sharpness

In certain situations, mathematicians prefer results which are not necessarily the simplest possible. One such situation involves statements dealing with an ordering of some sort. In this case, a sharper bound is generally preferred.

Definition 3.2 Let $\prec$ be a transitive relation on a set S, and let $x, y, z \in S$, such that $x \prec z$ and $y \prec z$. Then y is said to be a (strictly) *sharper bound* than x (on z) if $x \prec y$ and $y \not\prec x$.

For example, if it is true that for certain sets A, B, and C, we have both $A \subseteq B$ and $A \cup C \subseteq B$, then a mathematician would probably want to record the latter fact as a Theorem, and ignore the former – provided, that is, that $A \cup C$ does not turn out to be equal to A. The reason for this preference is that since containment is a partial ordering (or more specifically, a transitive relation), and since we always have $A \subseteq A \cup C$, then one can easily deduce the former from the latter. The reverse, of course, does not follow. Note that, should $A \cup C = A$, then the simplicity rule above should apply, meaning that we would prefer the former result.

In MATHsAiD, whenever a relation is introduced, the system tries to determine whether the relation satisfies the usual properties one looks for in relations; namely, reflexivity, symmetry, anti-symmetry, and transitivity. Should the relation be found to be transitive, but not an equivalence relation, then the default setting (adjustable by the user) is to treat this relation henceforth as a type of ordering, and determine future Theorems on the basis of sharpness, rather than simplicity. This is, however, subject to the condition (as mentioned above) that if two terms are equal, then preference is given to the simpler term.

The reason for having the aforementioned setting be adjustable, is that there are, not surprisingly, exceptions to the preference for sharper bounds. For example, given positive integers a, b, c, d such that $a < b$ and $c < d$, the conclusion that $a + c < b + d$ is considered to be a Theorem, despite the existence of sharper results, say, $a + c < b + c$, for example. Moreover, 'implication' is certainly a transitive (not an equivalence) relation, but in this case, one would tend to prefer the bounds reversed. For example, if one knew that $A \implies B$ and that $B \implies C$, then one would likely choose $A \implies C$ as a Theorem, over the other two statements. This choice would usually be made, without regard to the simplicity of the respective arguments, but rather to the fact that the weaker hypothesis produces a stronger, and therefore more useful, overall result.

3.5 Converses

Among the various types of Theorems, mathematicians tend to value equivalences more highly than most. So much so, in fact, that they are usually willing to overlook it, should one of the implications be deemed unworthy, on its own, to be called a Theorem. This is, of course, provided that its converse is judged to be able to stand on its own. Quite simply, if an implication and its converse are both true, then the two of them together are more useful than either on its own.

For this reason, in MATHsAiD, whenever an implication is known, either as an axiom or as a Theorem, then an attempt [19] is made to prove its converse. Should this attempt succeed, then the converse is added [20] as a Theorem, without regard to any of the preceding conditions.

4 Results

Thus far, we have tested MATHsAiD in set-theory, the positive integers (without induction – that will come later), and in group theory. In each setting, we have only explored the very basics, but we would expect this to be sufficient, in order to see how well our Theorem filters work. The Theorems identified by MATHsAiD from each area are included in the Appendix, along with the axioms provided to the system, and compared with the Theorems listed in various textbooks on each subject. Note that, due to the variations in style amongst different authors, we have likewise allowed for some variations in what constitutes a Theorem in the textbooks. Certainly, any statement that has a specific label (e.g., Lemma, Corollary, etc.) is counted. In addition, if an unlabelled statement is set apart in some way, according to some pattern established within the text, along with some proof of the statement, then it too is counted as a Theorem. However, exercises are not counted as Theorems, although we point them out, when they match either a MATHsAiD or another textbook's Theorem.

In addition to the Theorems found by MATHsAiD, we also include each Theorem (as described above) found in the surveyed textbooks, that could reasonably be expected to have been found by MATHsAiD, given the axioms [21] and definitions provided to MATHsAiD. It turns out that MATHsAiD can – in manual mode – find at least some of these extra Theorems; the users of MATHsAiD may provide their own hypotheses to the system, and thereby generate more Theorems (as, indeed, we have done). However, for the purposes of this study, we only give MATHsAiD credit for the Theorems it generates in automatic mode.

[19] Only a 'reasonable' attempt, that is. Most of the time, a converse will not, in fact, be true. Thus a balance should be struck, between exerting enough effort to prove at least the routine results, whilst not wasting too much time and effort in trying to prove something that isn't true.

[20] We would, of course, tend to prefer the equivalence as a Theorem, rather than either of the two individual implications. For this paper, however, we have preserved the results, including the order of discovery, just as MATHsAiD produced them.

[21] For the time being, some "routine" information must also be provided by the user; e.g., the type of predicate involved, be it a relation, an operation, etc. It is expected that this can be automated in the future.

It should be noted that the axioms provided to MATHsAiD, in some cases, differ somewhat from those listed in the textbooks in this survey. For our purposes, however, at least in the examples provided, this does not seem to be all that pertinent. The one exception is in the definition of intersection given in [9], which is marked in the corresponding table.

We would like to point out that the discrepancies between MATHsAiD and each of the textbooks seem to be quite in line with the discrepancies amongst the textbooks themselves. It is somewhat interesting to note that, when it comes to Theorems that have both a left- and a right-handed version (e.g., the distributivity laws), MATHsAiD sometimes produces both versions, and sometimes only one. As it happens, this same phenomenon also occurs within textbooks.

We provide a summary of the results in the following tables, but we encourage the readers to examine the individual Theorems recorded in the Appendix, and to judge for themselves whether each so-called Theorem is indeed justly named. In each table, we compare MATHsAiD to two textbooks, and we give two scores – a stringent score and a more generous score – for each source of Theorems. Both scores, as defined below, are computed in the same way, but by using slightly different data. In the stringent score, the source has to have the Theorem recorded almost verbatim, which shows up in the Appendix tables as a tick in the appropriate box. In the more generous score, we allow for more discrepancies, counting all marks in the boxes, except for T's, S's and C's[22] (see the notes after each table).

Our scoring system was chosen primarily because of its simplicity and its fairness. In particular, while it is certainly not fair to give credit to a given source for listing loads of so-called Theorems that no other source calls by that name, it likewise is unfair to the other sources, not to penalise the given source for having done so. Similarly, a source should be penalised for failing to recognise as a Theorem, any statement that both of the other sources identified as such. Note that, with this scoring system, the higher the score, the better. We would hope that the scores for MATHsAiD would compare favourably with the scores for the textbooks. That is to say, its score should be roughly in the range of the other two, to which it is being compared.

Each of the tables below has the following key:

s – stringent criteria

g – generous criteria

T(c) – Number of Theorems listed by this source, under criteria c

A(c) – Number of Theorems this source alone listed, under criteria c

M(c) – Number of Theorems this source alone failed to list, under criteria c

MAT – MATHsAiD

[22] In each of these cases, it does not seem fair to give MATHsAiD credit for having discovered – but discarded – the Theorems. However, in cases of G, for instance, the Theorems were filtered out precisely because of the sharpness criterion (which is adjustable by the user).

Definition 4.1 Given the notation above, the *score*, under criteria c, for each source is given by: Score(c) = (T(c) - 1.5A(c)) - M(c).

Set Theory

Source	T(s)	A(s)	M(s)	Score(s)	T(g)	A(g)	M(g)	Score(g)
[9]	34	4	5	**23**	40	4	6	**28**
[12]	28	4	11	**11**	31	4	15	**10**
MAT	44	13	4	**20.5**	45	6	3	**33**

Positive Integers

Source	T(s)	A(s)	M(s)	Score(s)	T(g)	A(g)	M(g)	Score(g)
[4]	19	3	1	**13.5**	24	3	1	**18.5**
[10]	11	0	6	**5**	15	0	7	**8**
MAT	21	6	2	**10**	23	1	0	**21.5**

Group Theory

Source	T(s)	A(s)	M(s)	Score(s)	T(g)	A(g)	M(g)	Score(g)
[1]	6	0	2	**4**	7	0	2	**5**
[8]	9	1	0	**7.5**	9	0	0	**9**
MAT	8	0	0	**8**	8	0	1	**7**

Note that in each of the six sets of scores, MATHsAiD either scores the highest, or second best.

5 Conclusions

We hoped to identify various criteria by which mathematicians determine which theorems to designate as Theorems, in recognition of the importance of this ability to separate the wheat from the chaff – not only for mathematicians, but potentially for artificial intelligence and machine learning systems as well. We then wanted to incorporate these criteria into an automated system, MATHsAiD, that would likewise discern between the two types of truths, hoping to produce results comparable to those produced by the human mathematical process. In order to judge our results, we compared the Theorems produced by MATHsAiD with Theorems listed in various mathematics textbooks.

We feel that, while there is still some room for improvement, not to mention some work yet to be done, our efforts can, on the whole, be considered successful. It is true that we have thus far only worked with fairly simple mathematical structures, and even then, only at an introductory level. However, these structures are quite different from one another, and yet the results are, arguably, at least, equally good in each area. Moreover, the fact that our methods are quite simple – yet effective – would seem to be rather significant.

Acknowledgement

The authors would like to thank the referees for their valuable comments and suggestions.

A Appendix

All the data contained here have been rewritten in standard mathematical notation, for the convenience of the reader. With one exception (the definition of intersection), the axioms/definitions provided to MATHsAiD are essentially the same as those found in the textbooks used for comparison.

A.1 Set Theory

The following are the set-theory axioms/definitions provided to MATHsAiD.

Axioms: Given that A and B are sets;	
1. $\forall x, x \notin \varnothing$	6. $A \cap B$ is a set
2. $\varnothing$ is a set	7. $\forall x, [(x \in A \text{ or } x \in B) \iff x \in A \cup B]$
3. $[\forall x, (x \in A \implies x \in B)] \iff A \subseteq B$	8. $A \cup B$ is a set
4. $A = B \iff [\forall x, (x \in A \iff x \in B)]$	9. $\forall x, [(x \in A \text{ and } x \notin B) \iff x \in A - B]$
5. $\forall x, [(x \in A \text{ and } x \in B) \iff x \in A \cap B]$	10. $A - B$ is a set

The following table contains the set-theory Theorems, as identified by MATHsAiD (MAT), [9], and [12]. Assume throughout that A, B, and C are sets.

Theorems	[9]	[12]	MAT
1. $A \subseteq A$.		✓	✓
2. $A \subseteq B$ and $B \subseteq A \implies B = A$.	✓	✓	✓
3. $A \subseteq B$ and $B \subseteq C \implies A \subseteq C$.	✓	✓	✓
4. $\varnothing \subseteq A$.	✓		✓
5. $A \subseteq B$ and $B \subseteq A \iff A = B$.			✓
6. $\forall x, x \notin A \iff \varnothing = A$.		✓	✓
7. $A \cap A = A$.	✓	✓	✓
8. $A \cap \varnothing = \varnothing$.		✓	✓
9. $B \cap A = A \cap B$.	✓	✓	✓
10. $(A \cap B) \cap C = A \cap (B \cap C)$.	✓	✓	✓
11. $A \cap B \subseteq A$.	✓	✓	✓
12. $A \cap B \subseteq B$.	✓		✓
13. $A \subseteq B \implies A \cap B = A$.	✓	✓	✓
14. $A \subseteq B \implies A \cap C \subseteq B \cap C$.	G		✓
15. $A \subseteq B \implies C \cap A \subseteq C \cap B$.	G		✓
16. $C \subseteq A$ and $C \subseteq B \implies C \subseteq A \cap B$.		D	✓
17. $A \cup A = A$.	✓	✓	✓
18. $A \cup \varnothing = A$.	✓	✓	✓
19. $B \cup A = A \cup B$.	✓	✓	✓

Theorems	[9]	[12]	MAT
20. $(A \cup B) \cup C = A \cup (B \cup C)$.	✓	✓	✓
21. $A \subseteq A \cup B$.	✓	✓	✓
22. $B \subseteq A \cup B$.	✓		✓
23. $A \subseteq B \implies A \cup B = B$.	✓	✓	✓
24. $A \subseteq B \implies A \cup C \subseteq B \cup C$.	G		✓
25. $A \subseteq B \implies C \cup A \subseteq C \cup B$.	G		✓
26. $A \subseteq C$ and $B \subseteq C \implies A \cup B \subseteq C$.		✓	✓
27. $A - A = \varnothing$.		✓	✓
28. $A - \varnothing = A$.	✓		✓
29. $\varnothing - A = \varnothing$.			✓
30. $B - (A - B) = B$.			✓
31. $(A - B) - A = \varnothing$.			✓
32. $A - (A - B) = A \cap B$.	DEF		✓
33. $(A - B) - B = A - B$.			✓
34. $(A - B) - C = A - (B \cup C)$.	✓		✓
35. $(A - B) \cup (A \cap C) = A - (B - C)$.	✓		✓
36. $A - B \subseteq A$.			✓
37. $A \subseteq B \implies A - B = \varnothing$.	✓		✓
38. $A \subseteq B \implies A - C \subseteq B - C$.	G		✓
39. $(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$.	✓	RHV	✓
40. $(A \cap B) \cup (A \cap C) = A \cap (B \cup C)$.	✓	RHV	✓
41. $(B \cap C) - A = B \cap (C - A)$.	✓		✓
42. $(B - A) \cup (C - A) = (B \cup C) - A$.	✓		✓
43. $A \cap B = A \implies A \subseteq B$.	✓	✓	✓
44. $A \cup B = B \implies A \subseteq B$.	✓	✓	✓
45. $A - (A \cap B) = A - B$.	✓	✓	
46. $A \cap (A - B) = A - B$.		✓	
47. $(A - B) \cup B = A \cup B$.	✓	✓	
48. $(A \cup B) - B = A - B$.		✓	
49. $(A \cap B) - B = \varnothing$.		✓	
50. $(A - B) \cap B = \varnothing$.		✓	
51. $(A - B) \cap (A - C) = A - (B \cup C)$.	✓	✓	S
52. $(A - B) \cup (A - C) = A - (B \cap C)$.	✓	✓	
53. $A \subseteq B \wedge C \subseteq D \implies A \cup C \subseteq B \cup D$.	✓		
54. $A \subseteq B \wedge C \subseteq D \implies A \cap C \subseteq B \cap D$.	✓		
55. $A \subseteq B \wedge C \subseteq D \implies A - D \subseteq B - C$.	✓		
56. $C \subseteq D \implies A - D \subseteq A - C$.	✓		

RHV – The right-hand version of this Theorem was listed in the book.

G – A more general version of this Theorem is listed in the text, and is given in the table (see Theorems 53, 54, and 55). Given the appropriate hypotheses, MATHsAiD actually finds the more general version, but discards it, because of the sharpness criterion.

D – The dual of this Theorem (namely, number 26) is in the book.

DEF – This is given as the definition of intersection.

S – A simpler equality for the right hand side was found by MATHsAiD (namely Theorem 34).

NOTE: We have not yet implemented within MATHsAiD a system whereby previously discovered Theorems are modified/deleted whenever subsuming Theorems are subsequently discovered. Hence, for example, Theorem 2 remains, even after Theorem 5 has been discovered. Also, as indicated in section 4, MATHsAiD sometimes reports both left- and right-handed versions of a Theorem, and at other times only reports one version. This depends on a variety of factors, particularly the input (including the axioms) provided to MATHsAiD, the previously discovered Theorems available, the nature/structure of the new Theorem itself, and the circumstances surrounding the situation. We tend to view this behaviour as a positive aspect of MATHsAiD, since this represents a phenomenon that can be observed amongst mathematics textbooks as well (see also Theorems 10-21 in A.2).

A.2 Positive Integers

The following are the axioms/definitions for the positive integers (essentially as found in [4] – without induction) provided to MATHsAiD. (NOTE: the last four axioms are the first author's own preference for stating that an operation is well-defined. See also the group axioms.)

Axioms: Given that $a, b, c \in \mathbb{N}$;

1. $\mathbb{N}$ is a set	11. $(a = b) \ \vee \ (\exists x \in \mathbb{N} \text{ s.t. } a + x = b) \ \vee \ (\exists y \in \mathbb{N} \text{ s.t. } a = b + y)$
2. $1 \in \mathbb{N}$	12. $a = b \implies [(\forall x \in \mathbb{N}, a + x \neq b) \ \wedge \ (\forall y \in \mathbb{N}, a \neq b + y)]$
3. $a + b \in \mathbb{N}$	13. $\exists x \in \mathbb{N} \text{ s.t. } a + x = b \implies [(a \neq b) \ \wedge \ (\forall y \in \mathbb{N}, a \neq b + y)]$
4. $a * b \in \mathbb{N}$	14. $\exists y \in \mathbb{N} \text{ s.t. } a = b + y \implies [(a \neq b) \ \wedge \ (\forall x \in \mathbb{N}, a + x \neq b)]$
5. $a + b = b + a$	15. $a < b \iff \exists x \in \mathbb{N} \text{ s.t. } a + x = b$
6. $a + (b + c) = (a + b) + c$	16. $a = b \implies c + a = c + b$
7. $a * b = b * a$	17. $a = b \implies a + c = b + c$
8. $a * (b * c) = (a * b) * c$	18. $a = b \implies c * a = c * b$
9. $a * (b + c) = (a * b) + (a * c)$	19. $a = b \implies a * c = b * c$
10. $a * 1 = a$	

The following table contains the positive-integer Theorems, as identified by MATHsAiD (MAT), [4], and [10]. Assume throughout that $a, b, c \in \mathbb{N}$.

Theorems	[4]	[10]	MAT
1. $a \nless a$.	E		✓
2. $a < b$ and $b < c \implies a < c$.	✓	✓	✓
3. $a = b \vee a < b \vee b < a$.	✓	✓	✓
4. $a = b \implies a \nless b$.	H	H	✓
5. $a = b \implies b \nless a$.	H	H	✓
6. $a < b \implies a \neq b$.	H	H	✓
7. $a < b \implies b \nless a$.	H	H	✓
8. $a < a + b$.		✓	✓
9. $b < a + b$.			✓
10. $a < b \implies a + c < b + c$.	✓	✓	✓
11. $a < b \implies c + a < c + b$.	✓		✓
12. $a < b \implies a * c < b * c$.	✓	✓	✓
13. $a < b \implies c * a < c * b$.	✓		✓
14. $c + a = c + b \implies a = b$.	✓		✓
15. $a + c = b + c \implies a = b$.	✓	✓	✓

Theorems	[4]	[10]	MAT
16. $c * a = c * b \implies a = b.$	✓		✓
17. $a * c = b * c \implies a = b.$	✓	✓	✓
18. $a + c < b + c \implies a < b.$	✓	✓	✓
19, $c + a < c + b \implies a < b.$	✓		✓
20. $a * c < b * c \implies a < b.$	✓	✓	✓
21. $c * a < c * b \implies a < b.$	✓		✓
22. $(b + c) * a = b * a + c * a.$	✓		
23. 1 is unique.	✓		C
24. $1 * a = a.$	✓		T
25. $a < b \ \wedge \ c < d \implies a + c < b + d.$	✓	✓	G
26. $a < b \ \wedge \ c < d \implies a * c < b * d.$	✓	✓	G

E – This is left as an exercise for the student.

H – The text states Theorem 3 in terms of "exactly one" of the statements is true, which we translate into "one and only one". These Theorems represent the "only one" portion of that Theorem. The MATHsAiD versions here (4, 5, 6, and 7) are a result of the way the axioms were provided to it.

C – This is filtered out because the cancellation Theorems (see 16 and 17) were found first.

T – This is filtered out because it is considered too trivial, given the commutativity axiom for multiplication.

G – This is a more general version of a Theorem that MATHsAiD got (see 10, 11, 12, and 13). It is filtered out because of the sharpness criterion.

A.3 Groups

The following are the axioms/definitions for groups provided to MATHsAiD.

Axioms: Given that $a, b, c \in G$;	
1. G is a set	7. $e * a = a$
2. $a * b \in G$	8. $a * e = a$
3. $a = b \implies c * a = c * b$	9. $a^{-1} \in G$
4. $a = b \implies a * c = b * c$	10. $a * a^{-1} = e$
5. $a * (b * c) = (a * b) * c$	11. $a^{-1} * a = e$
6. $e \in G$	

The following table contains the group-theory Theorems, as identified by MATHsAiD (MAT), [1], and [8]. Assume throughout that $a, b, c \in G$.

Theorems	[1]	[8]	MAT
1. $(a^{-1})^{-1} = a.$		✓	✓
2. $b * a = a \implies b = e.$	✓	✓	✓
3. $a * b = a \implies b = e.$	✓	✓	✓
4. $b * a = e \implies a^{-1} = b.$	✓	✓	✓
5. $a * b = e \implies a^{-1} = b.$	✓	✓	✓
6. $(a * b)^{-1} = b^{-1} * a^{-1}$		✓	✓
7. $c * a = c * b \implies a = b.$	✓	✓	✓
8. $a * c = b * c \implies a = b.$	✓	✓	✓
9. $c * c = c \implies c = e.$	E	✓	C

E – This is left as an exercise for the student.
C – This is filtered out because the cancellation Theorems (see 7 and 8) were found first.

NOTE: For those who were expecting to see the Theorems regarding unique solutions to the equations $a * x = b$ and $y * a = b$, we have not yet given MATHsAiD the ability to discover existence Theorems. We expect to do so in the future. As for the theorem $e^{-1} = e$, it is perhaps somewhat interesting to note that MATHsAiD indeed found this result. However, MATHsAiD agreed with the two textbooks, that this result should not be called a Theorem. All other results one might expect to see in group theory require more axioms/definitions than were presented to MATHsAiD (e.g., integer exponents, subgroups, homomorphisms, etc.)

References

[1] G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, 3rd Ed., Macmillan, New York, 1965.

[2] S. Colton, *Automated Theory Formation in Pure Mathematics*, Distinguished Dissertations Series, Springer-Verlag, London, 2002.

[3] J. Denzinger and S. Schulz, Recording and Analysing Knowledge-Based Distributed Deduction Processes, *Journal of Symbolic Computation*, Vol. 21, 523–541, 1996.

[4] C. W. Dodge, *Numbers & Mathematics,* 2nd Ed., Prindle, Weber & Schmidt, Inc., 1975.

[5] J. Draeger and S. Schulz, Improving the Performance of Automated Theorem Provers by Redundancy-free Lemmatization. In I. Russel and J. Kolen, editors, *Proc. of the 14th FLAIRS*, Key West, AAAI Press, 345–349, 2001.

[6] W M Farmer, J D Guttman, F J Thayer. Little Theories. *Automated Deduction–CADE-11, Lecture Notes in Computer Science,* 607:567-581. Springer Verlag, 1992.

[7] Y Gao, Automated Generation of Interesting Theorems, Masters Thesis, University of Miami, 2004.

[8] T. W. Hungerford, *Algebra*, Holt, Rinehart and Winston, New York, 1974.

[9] K. Kuratowski and A. Mostowski, *Set Theory*, Studies in logic and the foundations of mathematics, Vol. 86, Polish Scientific Publishers North Holland, Warszawa, 1968.

[10] E. Landau, *Foundations of Analysis,* The Arithmetic of Whole, Rational, Irrational and Complex Numbers, Chelsea, 1951.

[11] D. Lenat, *AM: An Artificial Intelligence approach to discovery in mathematics.* Unpublished PhD thesis, Stanford University, 1976.

[12] P. Suppes, *Axiomatic Set Theory*, Van Nostrand Reinhold, New York, 1960.